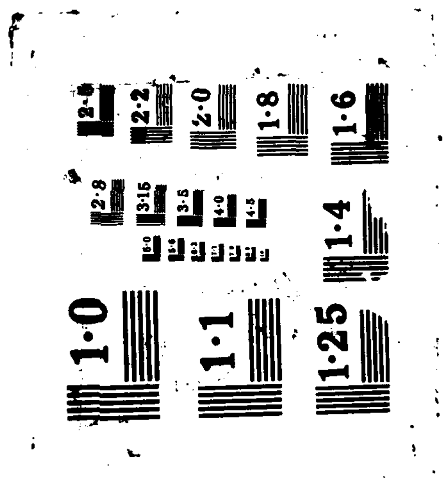
AD-A184 255 HETEROGENEOUS MULTI-TRUNKING QUEUEING SYSTEMS(U) GEORGE MASON UNIV FAIRFAX VA DEPT OF OPERATIONS RESEARCH AND APPLIED STATISTICS M J FISCHER ET AL SEP 87 1/1

UNCLASSIFIED GMU/22461/103 N00014-86-K-0029 F/G 12/4 NL



The State University
in Northern Virginia

(12)

DTIC FILE COPY

AD-A184 255

George Mason University

HETEROGENEOUS MULTI-TRUNKING QUEUEING SYSTEMS

by

Martin J. Fischer
Carl M. Harris

Report No. GMU/22461/103
September 1987

TECHNICAL REPORT

Office of Naval Research

Contract No. N0014-86-K-0029

HETEROGENEOUS MULTI-TRUNKING QUEUEING SYSTEMS

by

Martin J. Fischer
Carl M. Harris

Report No. GMU/22461/103
September 1987

Department of Operations Research and Applied Statistics
School of Information Technology and Engineering
George Mason University
Fairfax, VA 22030

Copy No. 9

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO. AD-A184

3. RECIPIENT'S CATALOG NUMBER 255

4. TITLE (and Subtitle)
Heterogeneous Multi-Trunking Queueing Systems

5. TYPE OF REPORT & PERIOD COVERED
Technical Report

6. PERFORMING ORG. REPORT NUMBER
GMU/22461/103

7. AUTHOR(s)
Martin J. Fischer
Carl M. Harris

8. CONTRACT OR GRANT NUMBER(s)
N00014-86-K-0029

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Operations Research and Applied Statistics
George Mason University, Fairfax, Va. 22030

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Task II
Project 4118150

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
800 North Quincy Street
Arlington, Va. 22217

12. REPORT DATE
September 1987

13. NUMBER OF PAGES
24

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

applied probability
computational probability
queueing systems
communication network
packet switching
probability
stochastic modeling
telecommunications

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A –

Abstract

In a packet switched network, several types of transmission may connect a pair of nodes of the network. The delay characteristics of each type of transmission may be significantly different and so a single queue to the node pair operating under a first-come, first-served rule may not be best. ~~In~~ this paper ~~we~~ develops both exact and approximate mathematical models that determine various system-level performance measures for a (K,N) scheme. Under this scheme arriving packets are placed in the queue for one type of transmission until the size reaches K, whereupon arriving packets are diverted to the other transmission system. When the number of packets at the first system drops to N(<K), arriving packets are again placed in the first queue.

I. Introduction

In a packet switched communication network, several transmission paths may exist between two switches in the network (see [1]). The delay characteristics of each path may be significantly different; for instance, one may be a satellite link (with its inherent propagation delay) and the other a 56kb/s terrestial link. Thus if packets from the same message are processed at the two links on a first-come, first-served basis, they will be received at the next node out of order. In order to alleviate this problem, a (K,N) scheme can be formulated for the operation of the system.

We consider the situation where there are two types of transmission and arriving packets are placed in the queue for the type 1 transmission system until the number of packets in that system is K, whereupon packets are sent to the second transmission system. Packets continue to be sent to the second queueing system until the number of packets at the first queueing system drops to N(<K), at which time arriving packets are again placed in the first system. The following figure depicts this scheme, which is known as the of heterogeneous, multi-trunking queueing system.

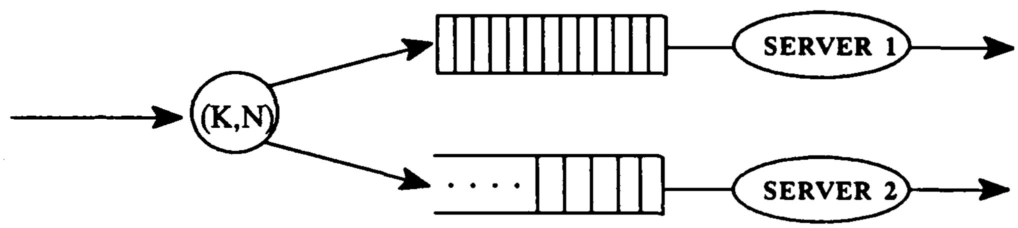


Figure 1. Heterogeneous Multi-Trunking Queueing System

We are considering a queueing system composed of two queues and a single arrival process of customers (packets). Arriving customers are placed into one or the other queue based on the number of customers at the first queue. Once assigned to a queue they remain there.

Queueing systems similar to this one have been considered before. In [2] Singh examined the single queue, first-come first-served version of the system. That is, customers are placed in service when a server is free based on their order of arrival. Another version of this system is the shortest queue problem where arriving customers are placed in the queueing system with the smallest number of customers (see [3] and [4] and the papers referenced therein).

In Section II we give a formal definition of the problem, our assumptions and a complete mathematical analysis of the system. The analysis required finding roots of an equation and the solution of a set of linear equations. In light of this, we developed

an approximation to the system-level performance that does not require these potentially complicated numerical problems. Section III documents the approximate model. Numerical comparisons between the exact and approximate models, as well as other numerical examples, are given in Section IV. Finally, Section V gives a summary of the paper.

II. Mathematical Performance Model

We are considering the system shown in Figure 1. We assume customers (packets) arrive in accordance with a Poisson process with rate λ. Let Q_i, i=1,2, be the steady-state number of customers in queueing system i. If Q_1=K, then arriving customers are placed in queueing system 2. This continues until Q_1=N (N<K), whereupon arriving customers are again placed in queueing system 1. Once placed in a system, a customer remains there. We assume that the length of time to serve a customer at server i is exponentially distributed with mean $1/\mu_i$, i=1,2, and the service-time distributions and arrival process are all independent of each other.

Under this rule the number of customers in system 1 is at most K, whereas there could be an infinite number at system 2. The value of N that is currently being considered in packet switched networks is N=K-1, so we first present an analysis of that case.

Case II.a: The (K,K-1) System

In this system, when the number of customers in system 1 reaches K, arriving customers are sent to system 2 until we have the first departure from system 1. Let us begin the analysis by defining

$$p_{ij} = \Pr\{Q_1=i,\ Q_2=j\} \quad (i=0,1,\ldots,K;\ j=0,1,\ldots). \tag{1}$$

Then the behavior at system 1 is given by the results of an M/M/1/K queueing system. Let $\rho_i = \lambda/\mu_i$, i=1,2; then from Gross and Harris [5], we have for i=0,1,...,K,

$$\Pr\{Q_1 = i\} = \begin{cases} \dfrac{\rho_1^i(1-\rho_1)}{1-\rho_1^{K+1}} & (\rho_1 \neq 1) \\ \dfrac{1}{K+1} & (\rho_1 = 1) \end{cases} . \tag{2}$$

The steady-state equations for p_{ij} are

$$\begin{cases} \lambda p_{00} = \mu_1 p_{10} + \mu_2 p_{01} \\ (\lambda+\mu_2) = \mu_1 p_{ij} + \mu_2 p_{0,j+1} \\ (\lambda+\mu_1+\mu_2)p_{ij} = \lambda p_{i-1,j} + \mu_1 p_{i+1,j} + \mu_2 p_{i,j+1} : i \leq K-1;\ j \geq 1 \\ (\lambda+\mu_1) = p_{K-1,0} + \mu_2 p_{K-1} \\ (\lambda+\mu_1+\mu_2)p_{K,j} = \lambda p_{K-1,j} + \lambda p_{K,j-1} + \mu_2 p_{K,j+1} : j \geq 1. \end{cases} \tag{3}$$

For $i=0,1,\ldots,K$, and $|z|\le 1$ define

$$P_i(z) = \sum_{j=0}^{\infty} p_{ij} z^j . \tag{4}$$

Then we have the following matrix equation for the $P_i(z)$:

$$\begin{bmatrix} a(z) & -\mu_1 z & 0 & & & \cdots & 0 \\ -\lambda z & b(z) & -\mu_1 z & 0 & & \cdots & 0 \\ 0 & -\lambda z & b(z) & -\mu_1 z & 0 & \cdots & 0 \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & \cdots & 0 & -\lambda z & b(z) & & -\mu_1 z \\ 0 & \cdots & 0 & 0 & -\lambda z & & b(z) \end{bmatrix} \begin{bmatrix} P_0(z) \\ P_1(z) \\ P_2(z) \\ \vdots \\ P_{K-1}(z) \\ P_K(z) \end{bmatrix}$$

$$= \mu_2 (z-1) \begin{bmatrix} p_{00} \\ p_{10} \\ p_{20} \\ \vdots \\ p_{K-1,0} \\ p_{K0} \end{bmatrix} , \tag{5}$$

where

$$\begin{cases} a(z) = (\lambda + \mu_2) z - \mu_2 \\ b(z) = (\lambda + \mu_1 + \mu_2) z - \mu_2 \\ c(z) = -\lambda z^2 + (\lambda + \mu_1 + \mu_2) z - \mu_2 \end{cases} . \tag{6}$$

Systems of equations like these appear frequently in these types of queueing problems and the solution technique is standard; for instance, see [6], [7], and [8].

Define the following recursions:

$$\begin{cases} E_{-1}(z) = 1 \\ E_0(z) = a(z) \\ E_i(z) = b(z)E_{i-1}(z) - \lambda\mu_1 z^2 E_{i-2}(z) & (i \leq K-1) \\ E_K(z) = c(z)E_{K-1}(z) - \lambda\mu_1 z^2 E_{K-2}(z) \end{cases} \tag{7}$$

and

$$\begin{cases} F_{-1}(z) = 1 \\ F_0(z) = c(z) \\ F_i(z) = b(z)F_{i-1}(z) - \lambda\mu_1 z^2 F_{i-2}(z) & (i \leq K-1) \\ F_K(z) = a(z)F_{K-1}(z) - \lambda\mu_1 z^2 F_{K-2}(z). \end{cases} \tag{8}$$

If A(z) is the (K+1)x(K+1) matrix multiplying the column vector of $\{P_i(z)\}$ on the left hand side of Equation (5), then

$$\det[A(z)] = E_K(z) = F_K(z). \tag{9}$$

Now let $A_i(z)$ be the matrix A(z) with the right-hand side of Equation (5) in the *ith* column, then, for i=0,1,...,K we have

$$P_i(z) = \frac{\det[A_i(z)]}{\det[A(z)]}, \tag{10}$$

where

$$\det[A_i(z)] = \mu_2(z-1)\left\{\sum_{k=0}^{i-1}(\lambda z)^{i-k} p_{k0} E_{k-1}(z) F_{K-1-i}(z) + \sum_{k=i}^{K}(\mu_1 z)^{k-i} p_{k0} E_{i-1}(z) F_{K-1-k}(z)\right\} \tag{11}$$

with the first sum set to zero when i=0. Thus the problem is solved once we find p_{i0}, i=0,...,K.

We need to develop K+1 equations in the unknowns $\{p_{i0}\ i=0,1,\ldots,K\}$. The first can be obtained from the normalizing condition $\left(\sum_{i=0}^{K} P_i(1)=1\right)$. It is straightforward to show, using this condition, that

$$\sum_{i=0}^{K} p_{i0} = 1-\rho_2 \frac{\rho_1^K(1-\rho_1)}{1-\rho_1^{K+1}} \tag{12}$$

and that the right-hand side of Equation (12) must be positive for Q_2 to be a well defined random variable. Since the quantity $\rho_1^K(1-\rho_1)/(1-\rho_1^{K+1})$ is the portion of time customers are arriving at system 2, ρ_2 times this quantity is the effective load on system 2, and this has to be less than 1 in order for p_{ij}, to be well defined. Thus the requirement that the right-hand side of equation (12) be positive makes intuitive sense. There are no existence conditions for Q_1, unless $K=\infty$.

It is easy to show that $\det[A(1)]\equiv 1$ and so z=1 is a root of the denominator of Equation (10). Using a similar analysis as in [6] or [8] one can show that there are K distinct roots, ξ_i, i=1,2,...,K, of $\det[A(z)]=0$ in (0,1). Since the numerator of Equation (10) must also vanish at these points, we have K additional equations given by

$$\sum_{k=0}^{K} (\lambda\xi_i)^{K-k} P_{k0} E_{k-1}(\xi_i) = 0 \qquad (i=1,2,\ldots,K). \tag{13}$$

Thus Equations (12) and (13) give us the required K+1 equations in the $\{p_{i0}\}$.

Summing Equation (5) we also have

$$P(z) = \sum_{i=0}^{K} P_i(z)$$

$$= 1-\rho_2 \frac{\rho_1^K(1-\rho_1)}{1-\rho_1^{K+1}} + \rho_2\, zP_K(z), \tag{14}$$

where $P_K(z)$ is given by Equations (10) and (11). From Equation (14) we have

$$\Pr\{Q_2=j\} = \begin{cases} 1-\rho_2 \dfrac{\rho_1^K(1-\rho_1)}{1-\rho_1^{K+1}} & (j=0) \\ \\ \rho_2 \, p_{K,j-1} & (j \geq 1) \end{cases} \quad . \tag{15}$$

Using Equations (2) and (14) one can compute the expected values of Q_1 and Q_2 and hence the expected value of the number of customers, $Q=Q_1+Q_2$, in the system. The average waiting time, E{W}, is then given via Little's Formula, $E\{W\} = \lambda^{-1}E\{Q\}$.

For the case when K=1 and N=0, we have

$$\begin{bmatrix} (\lambda+\mu_2)z-\mu_2 & -\mu_1 z \\ \\ -\lambda z & -\lambda z^2+(\lambda+\mu_1+\mu_2)z-\mu_2 \end{bmatrix} \begin{bmatrix} P_0(z) \\ \\ P_1(z) \end{bmatrix} = \mu_2(z-1) \begin{bmatrix} p_{00} \\ \\ p_{10} \end{bmatrix}, \tag{16}$$

$$\det[A(z)] = (z-1)\ (-\lambda^2 z^2 + \mu_2(\lambda+\mu_1+\mu_2)z-\mu_2^2). \tag{17}$$

So

$$P_0(z) = \frac{\mu_2\ [p_{00}(-\lambda z^2+(\lambda+\mu_1+\mu_2)z-\mu_2)+\mu_1 z p_{10}]}{-\lambda^2 z^2 + \mu_2(\lambda+\mu_1+\mu_2)z-\mu_2^2} \tag{18}$$

and

$$P_1(z) = \frac{\mu_2\ [\lambda z p_{00} + p_{10}((\lambda+\mu_2)z-\mu_2)]}{-\lambda^2 z^2 + \mu_2(\lambda+\mu_1+\mu_2)z-\mu_2^2} \tag{19}$$

where

$$p_{00} = \left(1-\rho_2 \frac{\rho_1}{1+\rho_1}\right) \frac{[1-\xi(1-\rho_2)]}{1-\xi} \tag{20}$$

and

$$p_{10} = \left(1-\rho_2 \frac{\rho_1}{1+\rho_1}\right) \frac{\rho_2 \xi}{1-\xi} \ , \tag{21}$$

with

$$\xi = \frac{\mu_2}{2\lambda}\left[(\lambda+\mu_1+\mu_2) - \sqrt{(\lambda+\mu_1+\mu_2)^2-4\lambda^2}\right] . \tag{22}$$

Even in this simple case no nice expression exists for $E\{Q_2\}$, so one is forced to differentiate $P(z)$.

Case II.b: The (K,N) System

For the case of a general N, we have to redefine p_{ij} to take into account which system is receiving arriving customers. When $N=K-1$ the value of Q_1 determines which system receives arriving customers. Begin here by letting I be an indicator variable such that

$$I = \begin{cases} 0 & \text{when arriving customers go to system 2} \\ 1 & \text{when arriving customers go to system 1} \end{cases}$$

and define

$$p^{l}_{ij} = \Pr\{Q_1=i,\ Q_2=j,\ I=l\}.$$

Now if $I=0$ then $p^{0}_{ij}=0$ for $i=0,1,\ldots,N$ and all j. If $I=1$ then $p^{1}_{Kj}=0$ for all j. As in case II.a one can write down the steady-state equations in p^{l}_{ij} and the resulting generating function equations are

$$A(z)\Pi(z) = \mu_2(z-1)\Pi_0 \tag{23}$$

where

$$A(z) = \begin{array}{c} \left[\begin{array}{ccccccccc}
a(z) & -\mu_1 z & 0 & & & & & \cdots & 0 \\
-\lambda z & b(z) & -\mu_1 z & 0 & & & & \cdots & 0 \\
0 & -\lambda z & b(z) & -\mu_1 z & & 0 & & \cdots & 0 \\
0 & & 0\ -\lambda z & b(z) & -\mu_1 z & 0 & & \cdots & -\mu_1 z \\
 & & 0 & -\lambda z & b(z) & -\mu_1 z & 0 & \cdots & 0 \\
 & & & -\lambda z & b(z) & -\mu_1 z & 0 & \cdots & 0 \\
 & & & 0 & c(z) & -\mu_1 z & 0 & \cdots & 0 \\
 & & & & \vdots & & & & c(z)-\mu_1 z \\
0 & \cdots & 0 & -\lambda z & 0 & \cdots & 0 & & c(z)
\end{array}\right] \begin{array}{l} \\ \\ \\ (N+1) \\ \\ (K) \\ \\ \\ \\ \end{array} \\ (K) \end{array} ,$$

$$
p(z) = \begin{bmatrix} P_0^1(z) \\ P_1^1(z) \\ \vdots \\ P_{K-1}^1(z) \\ P_{N+1}^0(z) \\ \vdots \\ P_K^0(z) \end{bmatrix}, \qquad P_0 = \begin{bmatrix} p_{00}^1 \\ p_{10}^1 \\ \vdots \\ p_{K-1,0}^1 \\ p_{N+1,0}^0 \\ \vdots \\ p_{K,0}^0 \end{bmatrix},
$$

and a(z), b(z), and c(z) are defined as before. We have annotated in which of the rows and columns of A(z) structural changes occurred. Again we have $\left(P_i^0(z) \equiv 0 \text{ for } i=0,1,\ldots,N\right)$

$$
P_i^0(z) = \mu_2(z-1)\frac{\det[A_i(z)]}{\det[A(z)]} \qquad (i=N+1,\ldots,K); \tag{24}
$$

where $A_i(z)$ is the matrix A(z) with P_0 in the K+i−N column. We start numbering the columns of A(z) with 0. For l=1 we have an analogous form for $P_i^1(z)$ for i=0,1,...,K−1 as that given for $P_i^0(z)$, except that $\det[A_i(z)]$ is the matrix A(z) with P_0 in the *ith* column. There is no simple recursion we can give for det[A(z)] or $\det[A_i(z)]$ as was done in the case of N=K−1. But the analysis is still the same; namely, one has to find the 2K−N−1 roots of det[A(z)] in (0,1), and set the numerator of one $P_i^1(z)$ to zero at these roots to find 2K−N−1 equations for $p_{i,0}^0$, i=N+1,...K and $p_{i,0}^1$, i=0,1,...,K−1.

The other equation is obtained from the normalization condition and we have

$$
\sum_{i=N+1}^{K} p_{i0}^0 + \sum_{i=0}^{K} p_{i0}^1 = 1-\rho_2 \ \Pr\{I=0\}. \tag{25}
$$

Thus we need only find an expression for Pr{I=0} to complete the solution.

In order to develop an expression for Pr{I=0} we use first-passage results from the standard M/M/1 queueing system (see [9]). Let τ_{ij} be the average first-passage time in system 1 from state i to state j; then

$$\Pr\{I=0\} = \frac{\tau_{KN}}{\tau_{KN} + \tau_{NK}} . \tag{26}$$

In system 1 the expected time to go from state K (Q_1=K) to N (Q_1=N) is the time for K–N departures, so that

$$\tau_{KN} = (K-N)/\mu_1. \tag{27}$$

Now to go from state N to state K for the first time in system 1 is the same as in a standard M/M/1 queue, and is given in [9] as

$$\tau_{KN} = \frac{\sum_{n=N}^{K-1} \frac{1-\rho_1^{n+1}}{\rho_1^n}}{\lambda(1-\rho_1)} . \tag{28}$$

Using all of these results we thus have

$$\Pr\{I=0\} = \frac{\rho_1(1-\rho_1)(K-N)}{\rho_1(1-\rho_1)(K-N) + \sum_{n=N}^{K-1} \frac{1-\rho_1^{n+1}}{\rho_1^n}} . \tag{29}$$

We may also study the behavior at system 1 in isolation. Using z=1 in the equations for $P_i^1(z)$ one gets the following equations for $p_i^1 = \Pr\{Q_1=i, I=1\}$:

$$\left\{ \begin{array}{ll} \lambda p_0^1 - \mu_1 p_1^1 = 0 & \\ -\lambda p_{i-1}^1 + (\lambda+\mu_1)\, p_i^1 - \mu_1 p_{i+1}^1 = 0 & (i=1,2,\dots,N-1) \\ -\lambda p_{N-1}^1 + (\lambda+\mu_1)\, p_N^1 - \mu_1 p_{N+1}^1 = \mu_1 p_{N+1}^0 & \\ -\lambda p_{i-1}^1 + (\lambda+\mu_1)\, p_i^1 - \mu_1 p_{i+1}^1 = 0 & (i=N,N+1,\dots,K-2) \\ -\lambda p_{K-2}^1 + (\lambda+\mu_1)\, p_{K-1}^1 = 0 & \end{array} \right. \tag{30}$$

and

$$p^0_{N+1} = p^0_{N+2} = ... = p^0_K = \rho_1 p^1_{K-1} . \qquad (31)$$

The solution to these equations is

$$p^0_i = \begin{cases} 0 & (i=0,1,...,N) \\ \dfrac{\rho_1^K(1-\rho_1)}{1-\rho_1^{K-N}} \; p^1_0 & (i=N+1,N+2,...,K) \end{cases}, \qquad (32)$$

and

$$p^1_i = \begin{cases} \rho_1^i \; p^i_0 & (i=0,1,...,N) \\ \dfrac{\rho_1^i - \rho_1^K}{1-\rho_1^{K-N}} \; p^1_0 & (i=N+1,N+2,...,K) \end{cases}, \qquad (33)$$

where

$$p^1_0 = \Pr\{I=0\} \; \frac{1-\rho_1^{K-N}}{\rho_1^K (K-N)(1-\rho_1)}, \qquad (34)$$

from which we have

$$\Pr\{Q_1=i\} = \begin{cases} \rho_1^i \; p^1_0 & (i=0,1,...,N) \\ \dfrac{\rho_1^i - \rho_1^{K+1}}{1-\rho_1^{K-N}} \; p^1_0 & (i=N+1,N+2,...,K) \end{cases}. \qquad (35)$$

For the case when $\rho_1=1$ we have

$$\Pr\{I=0\} = \frac{2}{3+K+N}$$

and

$$\Pr\{Q_1=i\} = \begin{cases} \dfrac{2}{3+K+N} & (i \le N) \\ \dfrac{2(K+1-i)}{(K-N)(3+K+N)} & (N+1 \le i \le K) \end{cases}. \qquad (36)$$

III. Approximate Behavior at System 2

In Section II we gave a complete analysis of the behavior of system 1, with results contained in Equations (32)-(36). The behavior of system 2 rests on finding the roots of an equation and then solving a system of equations. The results for system 1 are computationally simple enough; it would be nice to have similar simple expressions for system 2. Its behavior is similar to a standard M/M/1 queueing system where the arrival process of customers alternates between on and off times. Kuczura [10] presented the idea of an interrupted Poisson process (IPP) as a way of analyzing overflow streams in loss systems. Heffes [11] extended this work to a GI/M/1 queueing system where the arrival process is an interrupted Poisson process. His results will be used here to approximate the behavior at system 2. For an IPP it is assumed that the on and off time of the underlying Poisson process is described by independent exponentially distributed random variables. In general, this would not be true at system 2, except in the case of K=1, N=0. But the results of Heffes should provide an excellent approximation to the behavior of Q_2.

More formally, for system 2 we have a basic Poisson arrival process with rate λ. The input is turned on (off) for a period of time that is exponentially distributed with mean γ^{-1} (ω^{-1}). Following Heffes [11] the interarrival distribution at system 2 of customers is the mixture

$$A(t) = k_1\,(1-e^{-r_1 t}) + k_2(1-e^{-r_2 t}) \quad , \tag{37}$$

where for our system we have

$$\gamma = \mu_1/(K-N), \tag{38}$$

$$\omega = \frac{\lambda(1-\rho_1)}{\sum_{n=N}^{K-1} \frac{1-\rho_1^{n+1}}{\rho_1^n}} \quad , \tag{39}$$

$$r_1 = \frac{1}{2}\left\{\lambda+\omega+\gamma+\sqrt{(\lambda+\omega+\gamma)^2-4\lambda\omega}\right\}, \tag{40}$$

$$r_2 = \frac{1}{2}\left\{\lambda+\omega+\gamma-\sqrt{(\lambda+\omega+\gamma)^2-4\lambda\omega}\right\}, \tag{41}$$

and

$$k_1 = \frac{\lambda - r_2}{r_1 - r_2} \quad , \quad k_2 = 1 - k_1 . \tag{42}$$

Using standard results from the GI/M/1 queueing system (see [5]), we have the approximation to the number of customers at system 2 given by

$$E_A\{Q_2\} = \frac{\rho_2 \Pr\{I=0\}}{1-\theta} \tag{43}$$

where Pr{I=0} is given by Equation (29) and

$$\theta = \frac{-\eta_2 + \sqrt{\eta_2^2 - 4\eta_1\eta_3}}{2\eta_1} \tag{44}$$

with

$$\eta_1 = -\mu_2^2,$$

$$\eta_2 = r_2\,\mu_2 + r_1\,\mu_2 + \mu_2^2, \text{ and}$$

$$\eta_3 = -(k_1\, r_1\, \mu_2 + k_2\, r_2\, \mu_2 + r_1\, r_2).$$

We have just written down the expected value of Q_2; but using the results from the GI/M/1 queue one can give an expression for an approximation to the probability distribution of Q_2 as

$$\Pr\{Q_2 = i\} = \begin{cases} 1 - \rho_2 \Pr\{I=0\} & (j=0) \\ \rho_2 \Pr\{I=0\}(1-\theta)\theta^{\,j-1} & (j=1,2,\ldots). \end{cases} \tag{45}$$

One can use this approximation along with the exact results for system 1 (see Equation (35)) to give an easily computed approximation for the total number of customers in both systems. The average waiting time can then be found via Little's Formula. In the next section we compare the exact results with approximations.

IV. Numerical Examples

In this section we present some numerical examples using the results that were developed in Sections II and III. We investigate three areas:

a. Comparison of exact and approximate models;

b. Characterization of optimal K in the (K, K−1) system that minimizes expected number of customers in the system; and

c. Comparison of multi-trunking with first-come, first-served schemes.

Table 1 presents some comparisons of the exact results for the case K=2, N=1, and the IPP approximation of Section III. In the IPP case we are approximating the on and off periods of the arrival process to system 2 by exponential distributions with appropriate means. Since we can give exact values for the mean, the only facet of the approximation comes from the form of the underlying distribution. When K=1 and N=0, the on and off times are exponentially distributed with means μ_1^{-1} and λ^{-1} and the approximate and exact results do agree. For the case of N=K−1, the on time is again exponentially distributed with mean $1/\mu_1$ and so only the off time distribution is being approximated. Thus for the major cases of interest one would expect the approximation to be good. This is evident in Table 1.

$\mu_1 = \mu_2 = 1$		$\mu_1 = 5,\ \mu_2 = 1$		$\mu_1 = 10,\ \mu_2 = 1$	
λ	E{Q}	λ	E{Q}	λ	E{Q}
.3	.3682 .3680	2	.8073 .7944	2	.3082 .3068
.6	.8320 .8287	2.5	1.352 1.316	3	.6556 .6461
.9	1.432 1.415	3	2.467 2.367	4	1.414 1.372
1.2	2.402 2.346	3.5	6.213 5.863	5	4.168 3.974

TABLE 1. Comparisons of Exact and Approximate Models.

In Table 1 we have compared the results of the approximation with the exact results for the case K=2, N=1. The exact results are given as the upper entry in each box. For increasing arrival rates three comparisons were made; namely, μ_1/μ_2 = 1, 5, and 10. For each case considered we see that: the approximation underestimated the exact result; as the arrival rate increased the approximation got worse; and the ratio of the service rates was not a factor in the goodness of the approximation.

Since the approximation is underestimating the exact results, one would expect that the off time of the arrival process at system 2 would be better approximated by a distribution whose coefficient of variation is greater than one. We know of no simple results for the IPP with non-exponential on/off times. The best the approximation did was for the case λ =.3, $\mu_1 = \mu_2 = 1$, with a relative percent error of .054%; and the worst case was a 5.6% error for $\lambda = 3.5$, $\mu_1 = 5$, and $\mu_2 = 1$. One would expect this type of behavior because as λ increases, the load on system 2 increases and the approximation is being used to a greater extent. But even in the case of $\lambda = 3.5$, $\mu_1 = 5$, and $\mu_2 = 1$, the accuracy is quite good. The fact that there were not any great differences in the accuracy of the approximation as μ_1/μ_2 ranges from 1 to 10 is quite a surprise and perhaps follows from the fact that the mean off time (see Equation (39)) depends only on λ and ρ_1 and not on μ_1 and μ_2. All in all, we are quite satisfied with how well the IPP approximation does.

When cost is not a consideration, one might guess that the best value of K in a (K, K-1) system would be the value of K such that $K \cong \mu_1/\mu_2$ where $\mu_1 \geq \mu_2$, and one would select the integer closest to μ_1/μ_2 if the quotient were not an integer. Using the approximate model for Q_2 and the exact results for Q_1, we investigated three cases to try to determine if this were so. The results are given in Figures 2, 3, and 4. In each figure the values of $E\{Q_1\}$, $E\{Q_2\}$ and $E\{Q\} = E\{Q_1\} + E\{Q_2\}$ are given as a function of K. For all three figures one sees that the minimum of $E\{Q\}$ does not occur exactly at $K = \mu_1/\mu_2$ but is close. Figures 3 and 4 demonstrate that the optimal value of K does depend on λ as well as μ_1/μ_2. In the case of $\lambda = 4$ ($\mu_1/\mu_2=10$), the expected number of customers in the system was constant for $K \geq 5$. Whereas when $\lambda = 8$ the minimum occurred in the neighborhood of $K = 9$.

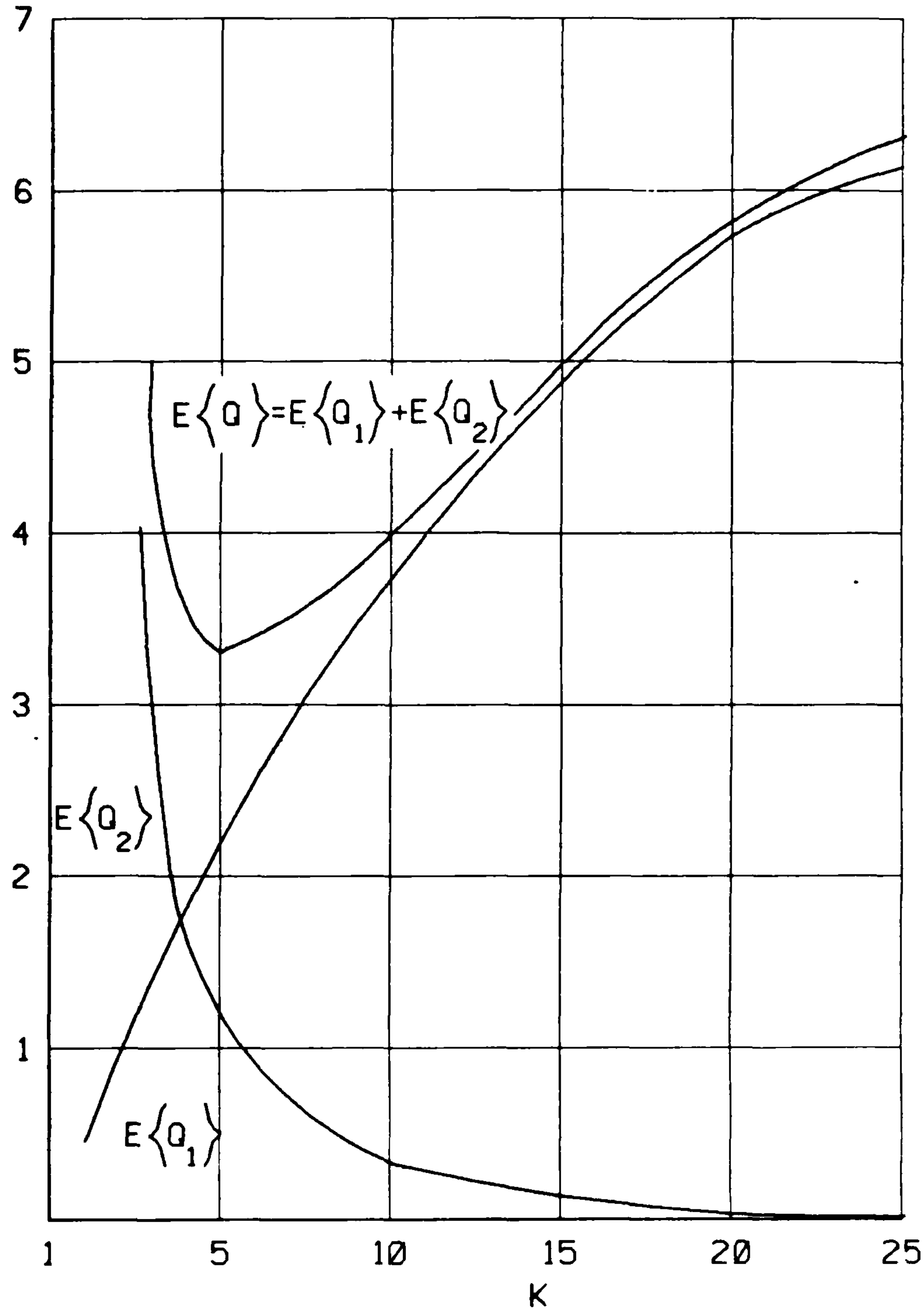


FIGURE 2. EXPECTED QUEUE LENGTHS AS A FUNCTION OF K FOR $\lambda=3.5$, $\mu_1=4$, AND $\mu_2=1$

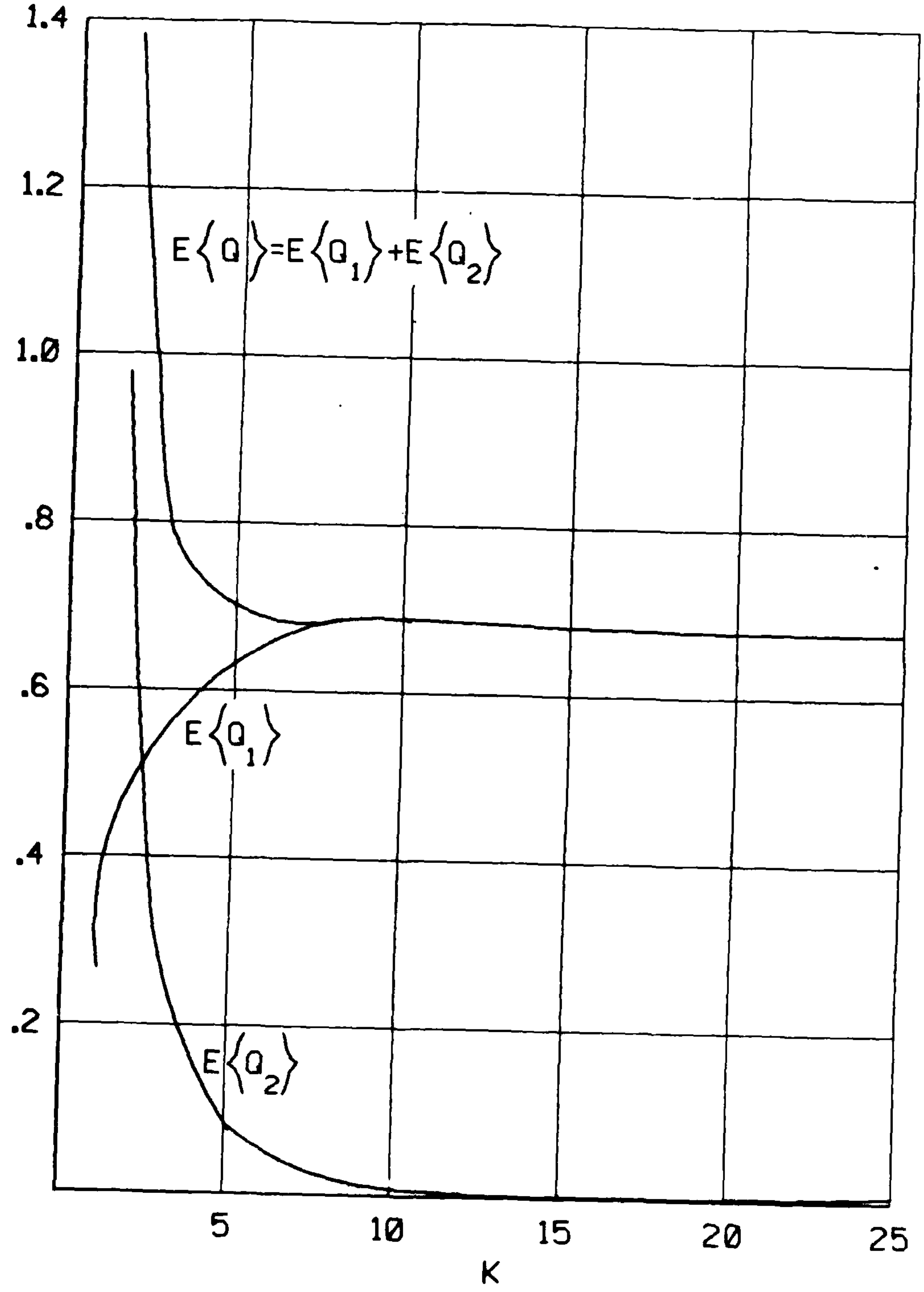


FIGURE 3. EXPECTED QUEUE LENGTHS AS A FUNCTION OF K FOR $\lambda=4$, $\mu_1=10$, AND $\mu_2=1$

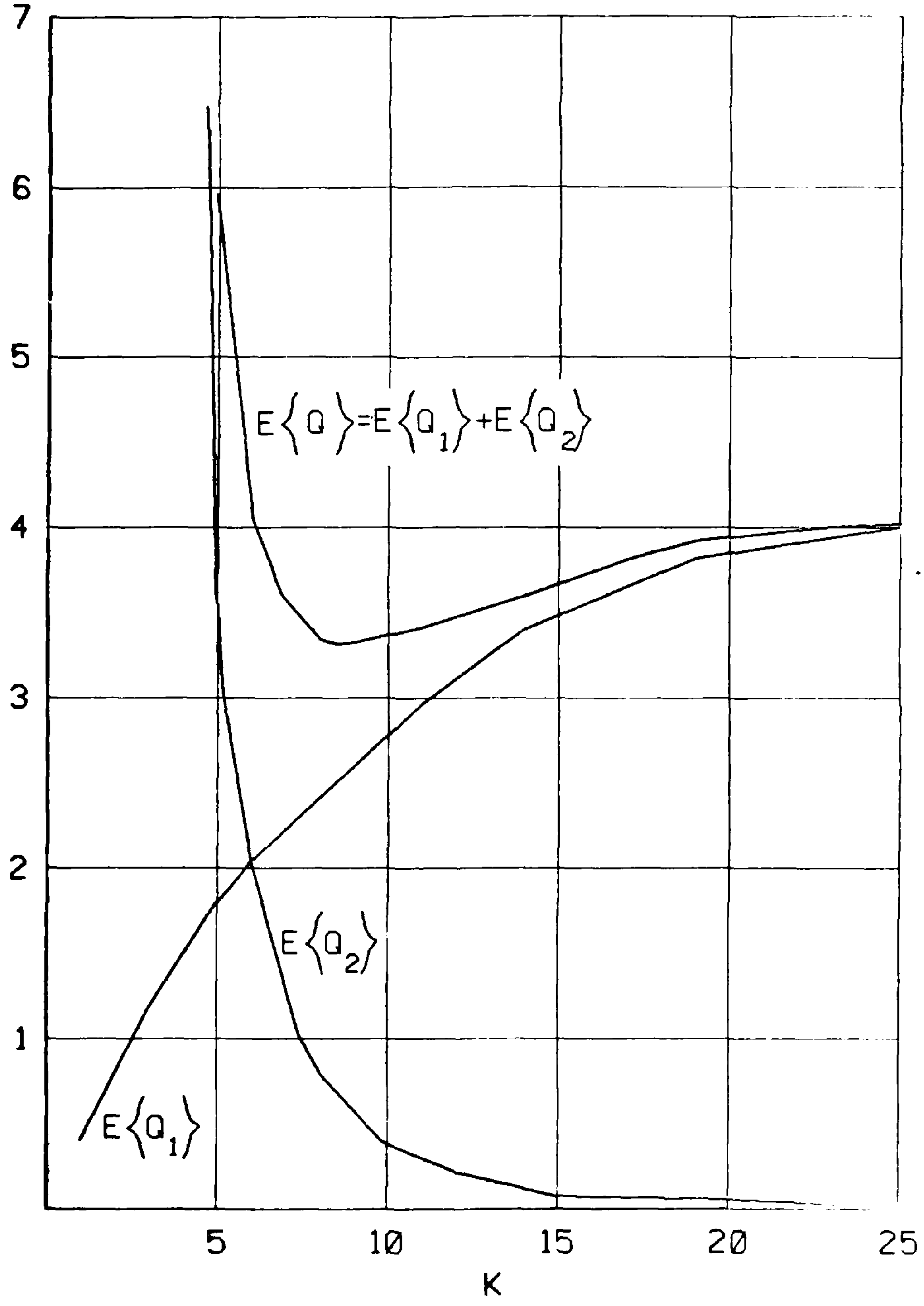


FIGURE 4. EXPECTED QUEUE LENGTHS AS A FUNCTION OF K FOR $\lambda=8$, $\mu_1=10$, AND $\mu_2=1$

At first glance one would expect that the (K, K–l) scheme would result in a smaller expected number of customers in the system than if one just formed one queue and placed customers in service as a server became free. The second system was studied by Singh in [2], and his results were used to compare with the multi-trunking scheme (K, K–l). This analysis is given in Tables 2, 3, and 4. In each of the tables, the total service rate $\mu_1 + \mu_2$ was fixed at 20 and μ_1 was varied. For each μ_1, we found the K^* which minimized the expected number of customers in the multi-trunking system, and we used the IPP approximation at system 2. For each table, one sees that the expected number of customers in the multi-trunking system is less than that under a single queue with first-come, first-served rule (Singh) only when μ_1/μ_2 is large. But, in general, the results of the two rules are relatively close for the cases we considered.

As we pointed out in Section I, the multi-trunking scheme tends to keep packets from the same message in the appropriate order of arrival at the receiving node. The main reason why a single queue with first-come, first-served rule (Singh) sometimes results in a shorter total queue is that there are times under the multi-trunking scheme where server 1 is idle and server 2 has customers requiring service.

Several other interesting facts can be seen from these tables. First, as μ_1 decreases, the expected number of customers increases for the multi-trunking scheme. Thus, if μ_1 ar.d μ_2 are relatively close in value, then one would not want to consider a multi-trunking type of scheme. For Singh's system, the behavior of the expected number of customers in the system is convex with a minimum occurring in the neighborhood of $\mu_1 = 12$.

μ_1	K	Multi-Trunking	Singh
19	≥ 12	1.111	1.705
18	7	1.234	1.538
17	5	1.338	1.436
16	4	1.418	1.373
15	3	1.487	1.333
14	3	1.534	1.311
13	2	1.587	1.303
12	2	1.588	1.303
11	2	1.616	1.315
10	2	1.664	1.331

TABLE 2. Expected Numbers of Customers – Multi–Trunking vs. Singh – $\lambda = 10$ and $\mu_1 + \mu_2 = 20$.

μ_1	K	Multi-Trunking	Singh
19	12	2.732	3.129
18	8	2.967	2.988
17	6	3.126	2.888
16	5	3.236	2.818
15	5	3.338	2.771
14	4	3.370	2.735
13	3	3.481	2.726
12	3	3.459	2.722
11	3	3.507	2.729
10	2	3.545	2.745

TABLE 3. Expected Numbers of Customers - Multi-Trunking vs. Singh - $\lambda = 14$ and $\mu_1 + \mu_2 = 20$.

μ_1	K	Multi–Trunking	Singh
19	19	9.721	9.847
18	14	9.737	9.678
17	11	9.881	9.636
16	9	10.051	9.567
15	8	10.188	9.517
14	7	10.356	9.484
13	6	10.557	9.464
12	6	10.676	9.457
11	5	10.793	9.460
10	5	10.953	9.474

TABLE 4. Expected Numbers of Customers – Multi–Trunking vs. Singh – $\lambda = 18$ and $\mu_1 + \mu_2 = 20$.

V. Summary

We have presented exact and approximate performance models for a multi-trunking scheme for packet switched networks. The exact solution was numerically complicated by the fact that one had to solve for roots of an equation and then solve a system of linear equations. The approximation required none of these complications and gave results that were in good agreement with the exact approach.

Using the approximate model, we made some comparisons with a single queue with first-come, first-served rule and found that the multi-trunking scheme yields a smaller expected number of customers only when the ratio of service rates is large.

Finally, we investigated the optimal value of K that minimized the expected number of customers in the (K, K-1) system. It was found to occur around the value of $K = \mu_1/\mu_2$ (with no costs considered). A very interesting problem to consider would be to determine whether the optimal value of K is changed under an imposed cost structure. For instance, if there is a significant cost to switch customer arrivals from system 1 to system 2, then one would expect the optimal value of K to vary significantly.

References

[1] M. Steenstrup, L. Denenberg and H. Mills, "Store and Forward Enhancement Algorithms Design," BBN Report No. 6360 (1987).

[2] V. P. Singh, "Two-Server Markovian Queues with Balking: Heterogeneous vs Homogeneous Server," *Oper. Res.* 18, 145–159 (1970).

[3] S. Halfin, "The Shortest Queue Problem," *J. Appl. Prob.*, 22, 865–878 (1985).

[4] C. Knessl, B. J. Mathowsky, Z. Schuss and C. Tier, "Two Parrallel Queues with Dynamic Routing," *IEEE Transactions on Comm.* Vol–Com–34, No. 12 (1986).

[5] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 2nd Edition, John Wiley & Sons, New York (1985).

[6] I. L. Mitrany and B. Avi–Itzhak, "A Many-Server Queue with Service Interruptions," *Oper. Res.* 16, 628–638 (1968).

[7] U. N. Bhat and M. J. Fischer, "Multichannel Queueing Systems with Heterogenous Classes of Arrivals," *Naval Research Logistics Quarterly*, 23, 271–282 (1976).

[8] B. Kraimeche and M. Schwartz, "Analysis of Traffic Access Control Strategies in Integrated Service Networks," *IEEE Transactions on Comm.*, Vol–Com–30, No. 10 (1985).

[9] J. W. Cohen, *The Single Server Queue*. Rev. Edition, North–Holland Publishing Co., Amsterdam, (1982).

[10] A. Kuczura, "The Interrupted Poisson Process As An Overflow Process," *B.S.T.J.*, Vol. 52, No. 3 (1973).

[11] H. Heffes, "Analysis of First–Come First–Served Queueing System with Peaked Inputs," *B.S.T.J.*, Vol. 52, No. 7 (1973).

DISTRIBUTION LIST

Copy No.	
1	Office of Naval Research 800 North Quincy Street Arlington, VA 22217 Attention: Scientific Officer, Statistics and Probability Mathematical Sciences Division
2	ONR Resident Representative Joseph Henry Building, Room 623 2100 Pennsylvania Avenue, N.W. Washington, DC 20037
3 - 8	Director, Naval Research Laboratory Washington, DC 20375 Attention: Code 2627
9 - 20	Defense Technical Information Center Building 5, Cameron Station Alexandria, VA 22314
21 - 29	C. M. Harris
30	GMU Office of Research

END

10-87

DTIC